## 5.5 Steamroller: An x86-64 Core Implemented in 28nm Bulk CMOS

Kevin Gillespie[1], Harry R. Fair III[1], Carson Henrion[2], Ravi Jotwani[3], Stephen Kosonocky[2], Robert S. Orefice[1], Donald A. Priore[1], Jonathan White[1], Kathryn Wilcox[1]

[1]AMD, Boxborough, MA, [2]AMD, Fort Collins, CO, [3]AMD, Austin, TX

The AMD two-core x86-64 CPU module, codenamed "Steamroller", contains 236 million transistors implemented in 28nm high-κ metal gate (HKMG) bulk CMOS using 12 levels of metal. It is designed to operate from 0.8 to 1.45V. The CPU module occupies 29.47 mm$^2$, which includes two independent integer cores, two instruction decode units and shared instruction fetch, floating-point, and 2MB 16-way L2 cache units (Fig. 5.5.7). Along with the second instruction decode unit, this design includes a larger shared 96KB 3-way instruction cache and a 10KB L2 branch target buffer for improved single-threaded performance and multi-threaded throughput compared to a previous 32nm AMD x86-64 CPU codenamed "Bulldozer" [1].

The module design contains 63 unique custom or compiled macros and 436,770 scan-able flip-flops. The design includes a new error-tolerant (ET) flop family to address the higher soft error rate (SER) in bulk CMOS than in SOI CMOS. A distributed power-gating technique is used throughout the module, including the gating of 4-way (512KB) sections within the L2 cache. Another addition to the design is the power-supply monitoring (PSM) circuits.

Design challenges included a new metallization compared to the 32nm process. The 28nm metal stack included more 1× metal layers than the previous design, which was closer to optimal for the GPU design on the APU SoC (Fig. 5.5.1). The Steamroller design used two threshold voltage devices ($LV_t$, $RV_t$) for the majority of the design, with the addition of an increased channel length on the $RV_t$ device. In addition, $HV_t$ devices were used specifically for the distribution of the power-header enable signals. The percentage of different device types for the 32nm design is similar to the 28nm Steamroller design (Fig. 5.5.2).

The flops used in this design were based on flops used in the previous generation [2]. In addition to the dual-clock soft-edge flop, a single-clock soft-edge flop was used as well. The advantage of the single-clock soft-edge flop was a reduction in routing overhead for the clock signals from the gater to the flops. As already mentioned, ET flops (Fig. 5.5.3) were also added to the design due to the move from SOI to bulk. ET flops comprise about 3% of the total core flops. This design showed a 5× reduction in error rate (as measured with neutron beam experiments) in 28nm bulk compared to the default flop. The additional redundancy is needed only on the slave latch due to the fact that approximately 90% of the flops are clock-gated. Additionally, a fault insertion flow was created to identify flops sensitive to SER. The sensitive flops were replaced with a footprint-compatible flop that had an improved failure-in-time (FIT) value on the slave node of about 20% compared to the default flops with little timing overhead. These improved SER flops comprise ~70% of the total flops in the core.

An improved resonant clocking design [3] was incorporated into Steamroller. In addition to the core clock, the L2 clock was also designed to resonate. In the frequency range of 3 to 4GHz, measured Cac savings (defined as $Cac=P_{dynamic}/V^2f$) were 150pF.

Similar to other processor designs, more synthesis and fewer custom macros were used in this design than in previous generations. Based on a total gate count, the percentage of synthesized gates has increased by about 4% from generation to generation on the past three implementations of the core.

Reduction of static and dynamic power consumption is critical in supporting thermally power-constrained products, enabling voltage boost states, as well as enhancing battery life for mobile products. To this end, the Steamroller core module focused significant design effort in these areas. In addition to the Steamroller core improvements, the core is instantiated in an SoC that makes use of an adaptive clocking system to improve power efficiency [4].

Rather than use a power-gating ring as used on previous designs [5], an integrated power-gating solution is implemented that reduces core leakage by 90% when gated at a 5% area penalty. A programmable state machine controls in-rush current during initial power-up of the gated supply by sequencing various amounts of header devices (Fig. 5.5.4). To ease the test burden, a serial status chain validates the header-enable signal is void of stuck-at faults.

Initial estimates projected the second-level cache to contribute as much as 27% of the total core leakage power. Implementation of a traditional retention mode bitcell sleep solution [6] is challenging due to the latency sensitivity of L2 accesses and the lack of advance notice to wake the bitcells. Additionally, bitcell sleep does not reduce active power, which is a substantial component of total power in an L2 with high access rates. A way-based power-gating scheme (Fig. 5.5.5) was implemented to address the power challenges specific to a large and active L2. The 2MB L2 cache is constructed from eight 256KB banks. Each bank is 16-way set associative. Within each bank, groups of 4 ways called "way domains" (WD) are individually power-gated. To reduce both leakage and active power, only the desired numbers of WD are dynamically enabled. This feature is especially valuable for periods of frequent C6 usage in which the cache is not fully utilized. This results in lower core energy per operation for applications such as Blu-ray playback that do not require maximum performance. Power gating by index would require an L2 flush to increase the cache size; however, power gating by way requires only that the tags of the newly powered ways be initialized.

In-rush current must be carefully managed when dynamically increasing the cache size so state is retained in the enabled ways (especially during operation near the bitcell $V_{min}$). To limit in-rush current, each WD is controlled by four wake signals and one run signal tied to progressively larger headers. The L2 has its own programmable state machine similar to the core solution. Additionally, each WD is divided into two power regions that have independent wake signals. This solution allows way-based power gating to function in down-cache scenarios in which banks are disabled. Each data macro contains two power domains of 4 ways each and a central power domain that is independent of the way gating.

Each tag macro contains 4 ways of tag information in one domain and a central power domain. Simulations show a reduction in total core power of 5% leakage and 0.5% active per gated WD.

Each Steamroller core contains ten PSM circuits distributed within each core pair to allow high-speed monitoring and digitization of the power supply grid during a running workload. The PSM block consists of a 32-stage differential ring oscillator (RO) that is reset at the start of every core clock (cclk) cycle (Fig. 5.5.6). At the end of each clock cycle, the state of the RO is captured and encoded with a priority encoder to generate the least-significant 6b of its digital output. The upper 8b are generated by dual binary counters offset by 180 degrees, which count the number of RO oscillations. On the next cycle, the RO state information is used to select automatically the binary counter that is fully settled to avoid propagating invalid data to the downstream logic. PSM configuration registers are accessed by a test and functional serial interface (JTAG/SPMI), while real-time streaming supply data can be transmitted onto a debug bus and stored in a dedicated L2 cache bank for debug and timing analysis.

Instantaneous measurement of the power supply for the two Steamroller core pairs for a virus workload has been measured (Fig. 5.5.6). The PSM also contains a clock divider for down-sampling and additional functions for calculating the minimum, maximum, average approximation, and a voltage-crossing alarm during a specified polling period.

*Acknowledgements:*
The authors thank David Akeson, Michael Bates, Steven Bakke, Tom Burd, Rob Dupcak, Ross McCoy, Richard McGowen, Spence Oliver, Jeshuah Sniderman, and all the members of the Steamroller physical design team for their contributions.

*References:*
[1] T. Fischer, *et al.*, "Design Solutions for the Bulldozer 32nm SOI 2-Core processor module in an 8-Core CPU," *ISSCC Dig. Tech. Papers*, pp. 78-80, 2011.
[2] S. Dillen, *et al.*, "Design and Implementation of Soft-Edge Flip-Flops for x86-64 AMD Microprocessor Modules," *IEEE Custom Integrated Circuits Conf.*, pp. 1-4, 2012.
[3] V. Sathe, *et al.*, "Resonant Clock Design for a Power-Efficient High-Volume x86-64 Microprocessor," *ISSCC Dig. Tech. Papers,* pp. 68-70, 2012.
[4] A. Grenat, *et al.*, "Adaptive Clocking System for Improved Power Efficiency in a 28nm x86-64 Microprocessor," to appear in *ISSCC Dig. Tech Papers*, 2014.
[5] R. Jotwani, *et al.*, "An x86-64 Core Implemented in 32nm SOI CMOS," *ISSCC Dig. Tech. Papers*, pp. 106-107, 2010.
[6] Y. Wang, *et al.*, "A 4.0 GHz 291Mb Voltage-Scalable SRAM Design in a 32nm High-κ + Metal-Gate CMOS Technology with Integrated Power Management," *ISSCC Dig. Tech. Papers*, pp. 456-457, 2009.

 978-1-4799-0920-9/14/$31.00 

| Metal Stack | 32nm | 28nm |
|---|---|---|
| 1x | 3 | 6 |
| 1.25x | 2 | 0 |
| 2x | 3 | 3 |
| 4x | 1 | 1 |
| 16x | 2 | 2 |

| Transistor (normalized to 32nm HVT leakage) | 32nm | 28nm |
|---|---|---|
| HVT | 1.00 | 0.53 |
| RVT | 5.33 | 3.67 |
| LVT | 16.67 | 19.80 |

Figure 5.5.1: Metal stack and transistor comparison from 32nm to 28nm.

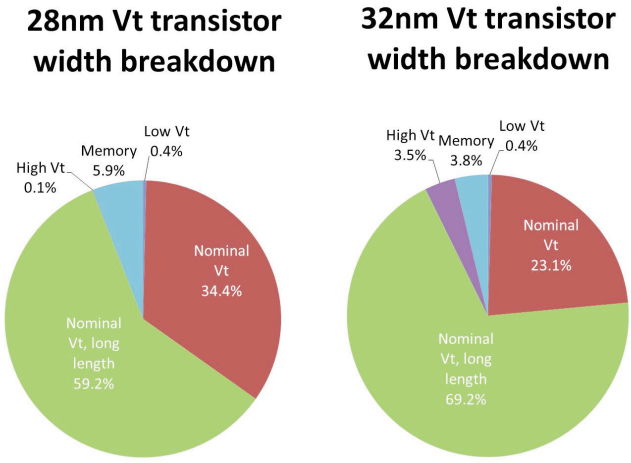


Figure 5.5.2: Transistor type breakdown.

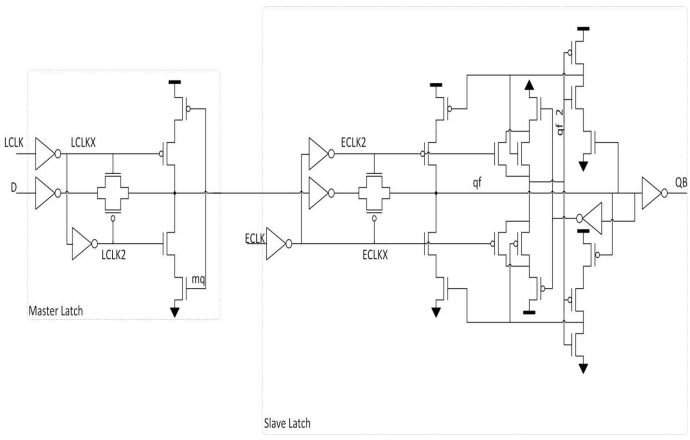


Figure 5.5.3: Error-tolerant flop.

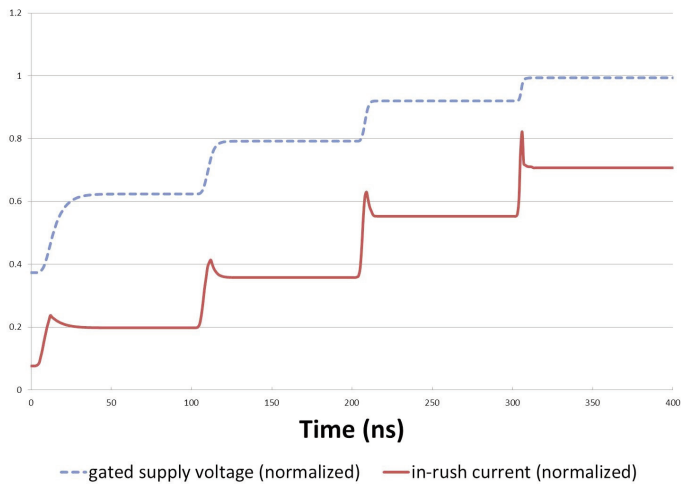


Figure 5.5.4: In-rush example based on one possible power-header state machine configuration.

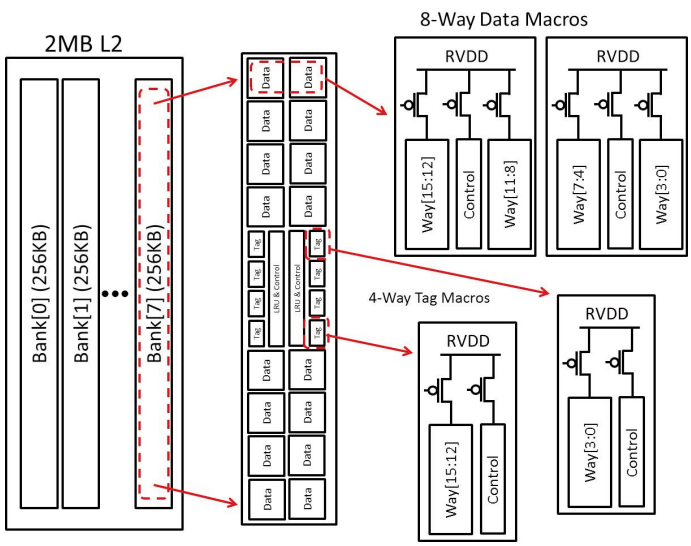


Figure 5.5.5: L2 way power gating.

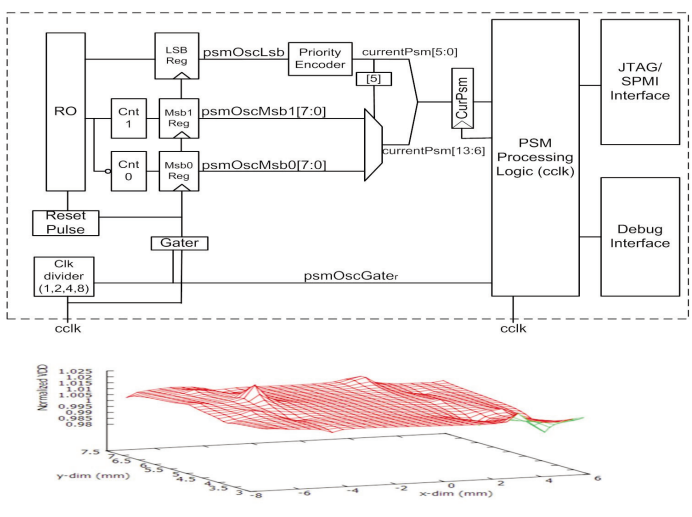


Figure 5.5.6: PSM logic and instantaneous measurement.

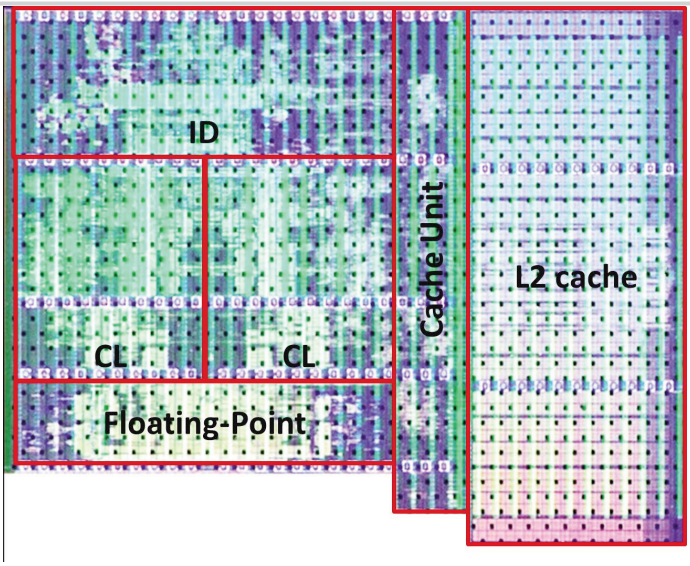


Figure 5.5.7: Die plot.


978-1-4799-0920-9/14/$31.00 